Overview of ICR Workspace

Thomas L. Casavant, Univ of Iowa/Holden

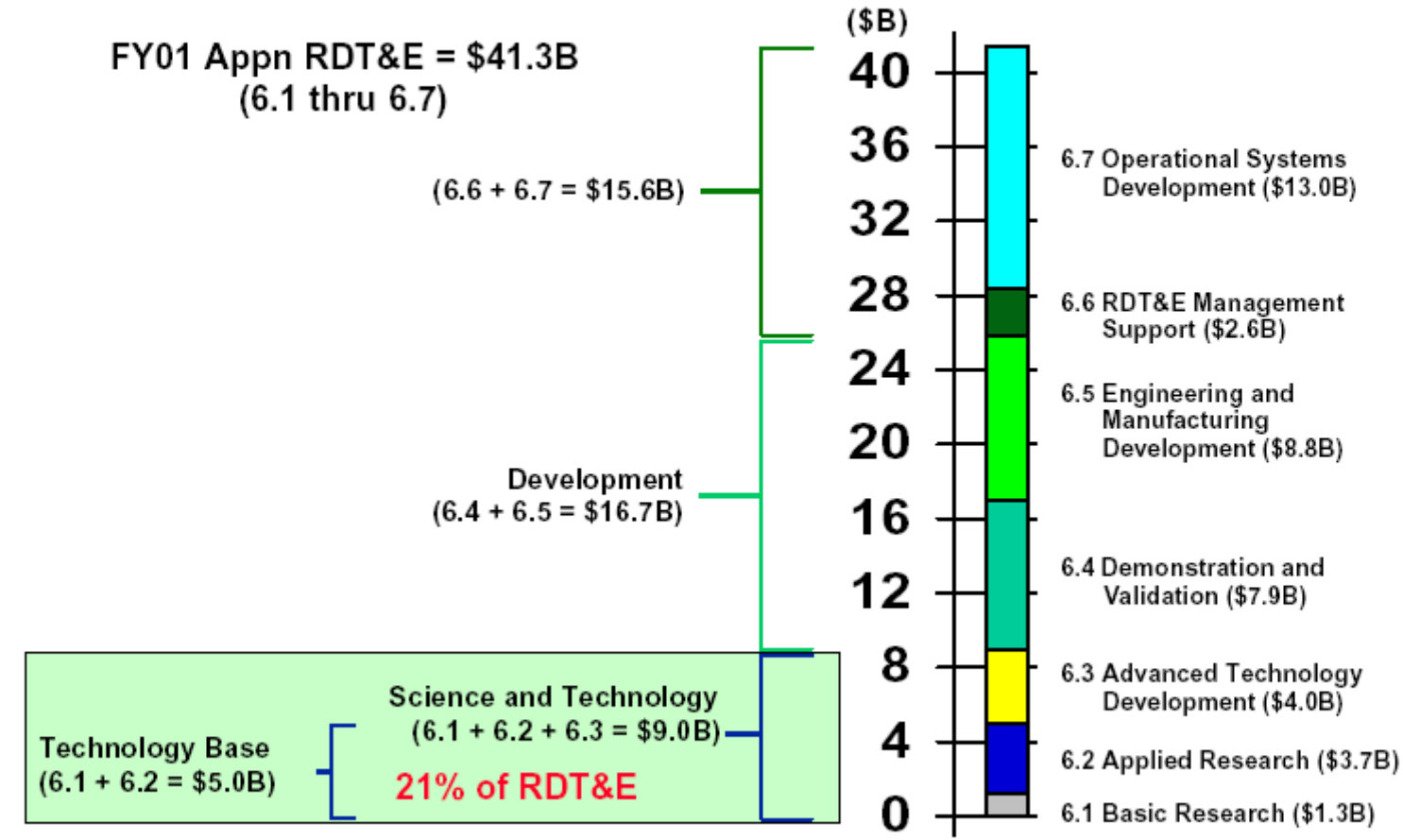
SP-SLWG Face to Face meeting

November 7-8, 2004

ICR Workspace at a Glance
- Overview::Definition

- The Integrative Cancer Research Workspace is producing modular and interoperable tools and interfaces that provide for integration between biomedical informatics applications and data. This will ultimately enable translational and integrative research by providing for the integration of clinical and basic research data. The Workspace is developing a software-engineered, well-documented and validated biomedical informatics toolset for use throughout the research community.
- DoD/DARPA-speak
  - I see this as a classic 6.1-6.2 charter, whereas CTMS is more 6.6

FY01 Appropriated RDT&E



ICR Workspace at a Glance - Overview::Composition

- Workspace Lead: Juli Klemm
- SP-SLWG Liaison – Michael Ochs- Fox Chase
- 6 ICR SIGs (8 proposed in February)
- ICR Membership
  - 152 on email alias
  - 69 attended 1st F2F meeting in Bethesda, August 25-26, 2004
    - 32 Institutions, plus
    - 14 from BAH and/or NCICB/SAIC
    - 9 from NCICB
    - 1 Oracle, 1 Alpha Gamma Technologies
- 21 ICR Development Projects
  - 19 Silver
  - 2 Gold reference implementations
- July/August caBIG program update highlighted ICR

ICR Workspace at a Glance
- Overview::Funded Participants

16 Developers:	7 Adopters:	4 Participants:
Columbia	Wistar	Vanderbilt -- Ingram
UNC Lineberger	New York University	U. of Michigan
UC San Francisco	Penn – Abrahamson	Prentis – Karmanos
Georgetown -- Lombardi	Memorial Sloan Kettering	Northwestern -- Lurie
Burnham Institute	Georgetown	
Wash U. -- Siteman	Oregon Health	
NCI – Center for Cancer Reseach	U. of South Florida -- Moffitt	
Cold Spring Harbor		
U. of Chicago		
Thomas Jefferson – Kimmel		
Memorial Sloan Kettering		
Fox Chase		
Dartmouth		
Duke		
University of Iowa -- Holden		
MIT/Broad		

ICR Workspace at a Glance - Overview::Liaisons

Tissue Banks and Pathology	Wash U Mark Watson
Clinical Trials Management	Penn David Fenstermacher
Architecture	Duke Patrick McConnell
Vocabularies and Common Data Elements	Vanderbilt Mary Edgerton
Data Sharing and Intellectual Capital	UI/Holden Tom Casavant/Terry Braun
Training	Institute for Cancer Prevention Edith Zang
Strategic Planning	Fox Chase Michael Ochs

ICR Workspace at a Glance
- Overview::Composition::6 SIGs

SIG (#projects, #members, #on a typical conference call)

1. Data Analysis and Statistical Tools (5, 52, 15-20)
   - Lead: Chris Kingsley, UCSF
2. Genome Annotation (6, 41, 10-15)
   - Lead: Craig Street, Penn
3. Microarray Repositories (3, 47, 15)
   - Lead: Julie Zhu, Northwestern
4. Pathways (3, 35, 10)
   - Lead: Shannon McWeeney, OHSU
5. Proteomics (3, 43, 10)
   - Lead: Sinoula Apostolou, Fox Chase
6. Translational (1, 40, 10)
   - Lead: Terry Braun, U Iowa/Holden

ICR Workspace at a Glance
- Overview::Brief Chronology Since Kickoff

February	- Kickoff Meeting - Special Interest Groups Defined
April	- First Workspace Teleconference - Juli Klemm starts as Workspace Lead - Begin conversations with Centers about projects
May	- Special Interest Group meetings begin - Continue project definitions - Begin Developer-Adopter pairings - Liaisons to other Workspaces identified - SOW template established
June	- Begin drafting SOWs - Project presentations begin in SIGs - Finalize Developer-Adopter pairings
July	- ICR featured in caBIG Program Update - Begin identification of candidate grid reference implementations - Continue drafting SOWs
August	- ICR face-to-face meeting - First Workspace Participant task orders issued
September	- Begin cost negotiations with Centers - First RFP issued (Duke) - Cancer Center leads for SIGs established
October	- Six more RFPs issued - First projects started: NCI-CCR's GoMiner and NCI-60 projects - VCDE liaisons established for SIGs - Developer/Adopter teams begin teleconferences, scheduling face-to-face meetings - SIGs begin discussion of exchange standards; Gene CDE focus group established - Begin UML-based CDE creation training

ICR Workspace at a Glance
- **Overview::**Meeting Schedule

- General Workspace Meetings
  - 2nd Wednesday of each month, 2:00pm - 3:00pm
- Data Analysis & Statistical Tools SIG
  - 1st Friday of each month, 2:00pm - 3:00pm
- Genome Annotation SIG
  - 1st Thursday of each month, 3:00pm - 4:00pm
- Informatics for Proteomics SIG
  - 2nd Monday of each month, 2:00pm - 3:00pm
- Microarray Repositories SIG
  - 1st Wednesday of each month, 2:00pm - 3:00pm
- Pathways Tools SIG
  - 1st Tuesday of each month, 1:00pm - 2:00pm
- Translational Tools SIG
  - 1st Monday of each month, 1:00pm - 2:00pm

SIG 1:
Data Analysis and Statistical Tools
(5, 52, 15-20)

The mission of the Data Analysis Special Interest Group of the ICR Workspace is to serve the needs of key categories of end users - experimentalists and data analysts - by provision of interoperable tools and associated standards, documentation, and training.

- Data analysis was identified as one of the most substantial needs across the Cancer Centers in the formal assessment that preceded initiation of the caBIG development workspaces.
- Much of the need stems from the increased complexity and volume of data sets resulting from high-throughput measurement technologies.

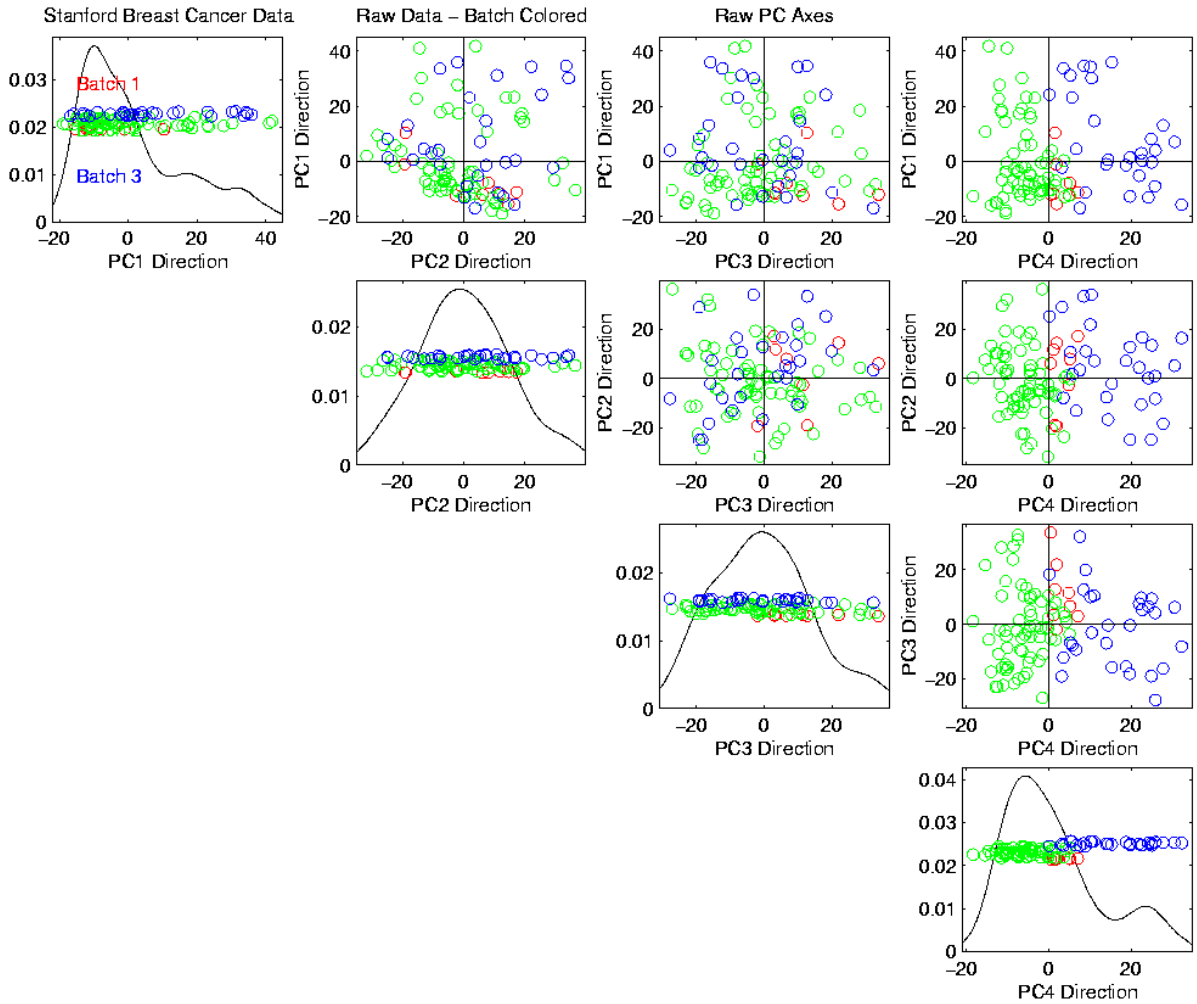
Distance-Weighted Discrimination (DWD)

Developer Center -- POC	UNC Lineberger – Steve Marron
Adopter Center -- POC	Wistar – Louise Showe
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

DWD Project Objectives

- Create a Risk Management Matrix for the project
- In collaboration with Adopter, create a use case document for DWD
- Develop Functional Requirements and Design Specification document, in collaboration with the Adopter Center and the cross-cutting Workspaces
- Document a Test Approach that ensures requirements are met
- Perform preliminary evaluation of Adopter's data
- Code caBIG-compliant DWD
- Create visual diagnostics for DWD
- Deploy caBIG-compliant DWD to Adopter site
- Execute on Test Approach

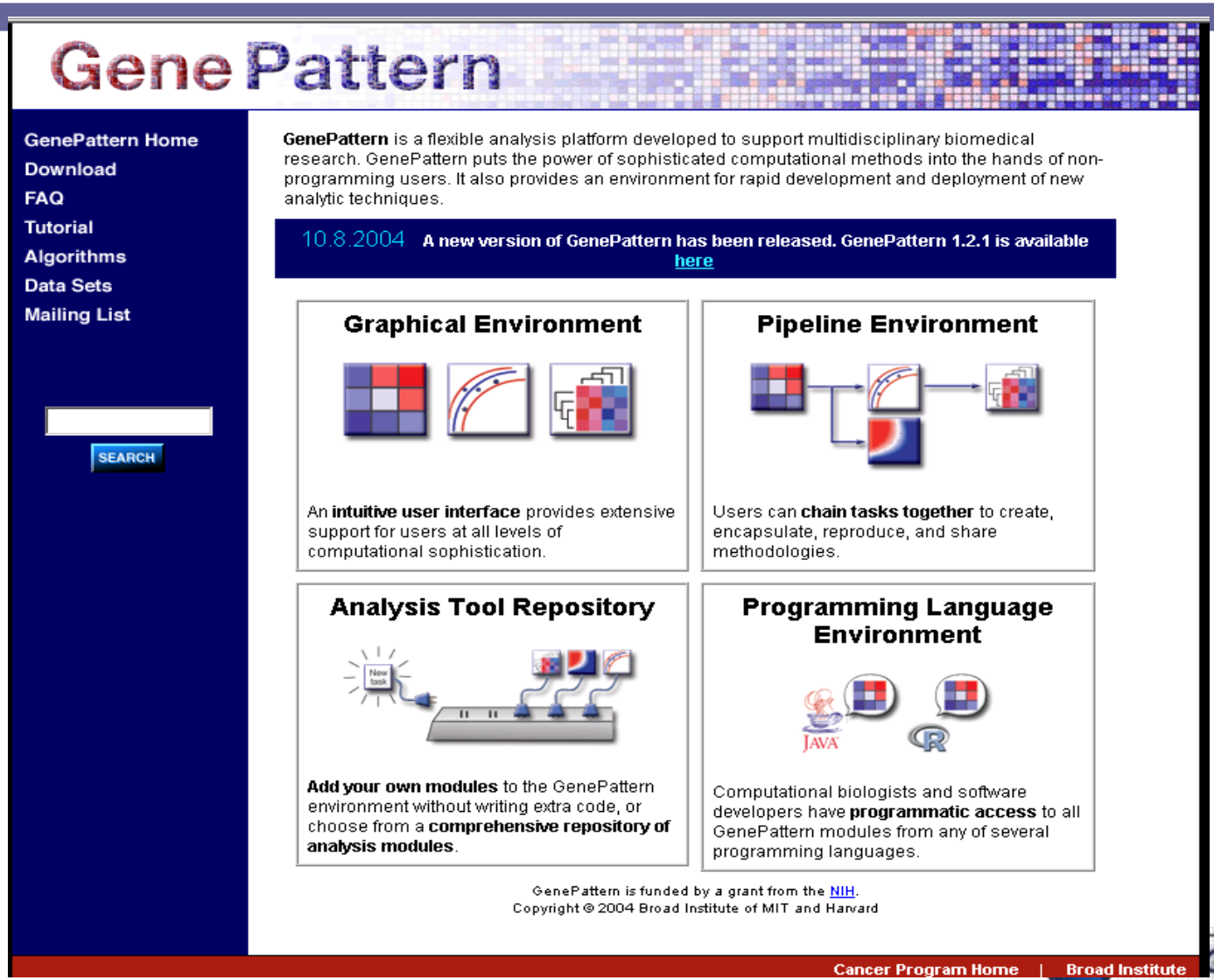
Source Batch Adj: Batch Colors



GenePattern

Developer Center -- POC	MIT/Broad – Jill Mesirov
Adopter Center -- POC	NYU -- Judith Goldberg
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

GenePattern



Magellan

Developer Center -- POC	UCSF – Ajay Jain
Adopter Center -- POC	Penn -- David Fenstermacher
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

Magellan Project Objectives

- Develop Functional Requirements and Design Specification documents, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code to achieve the following milestones:
  - Interoperability between Magellan and caBIO
  - Interoperability between Magellan and caArray
- Execute on Test Approach
- Deploy system to Adopter site

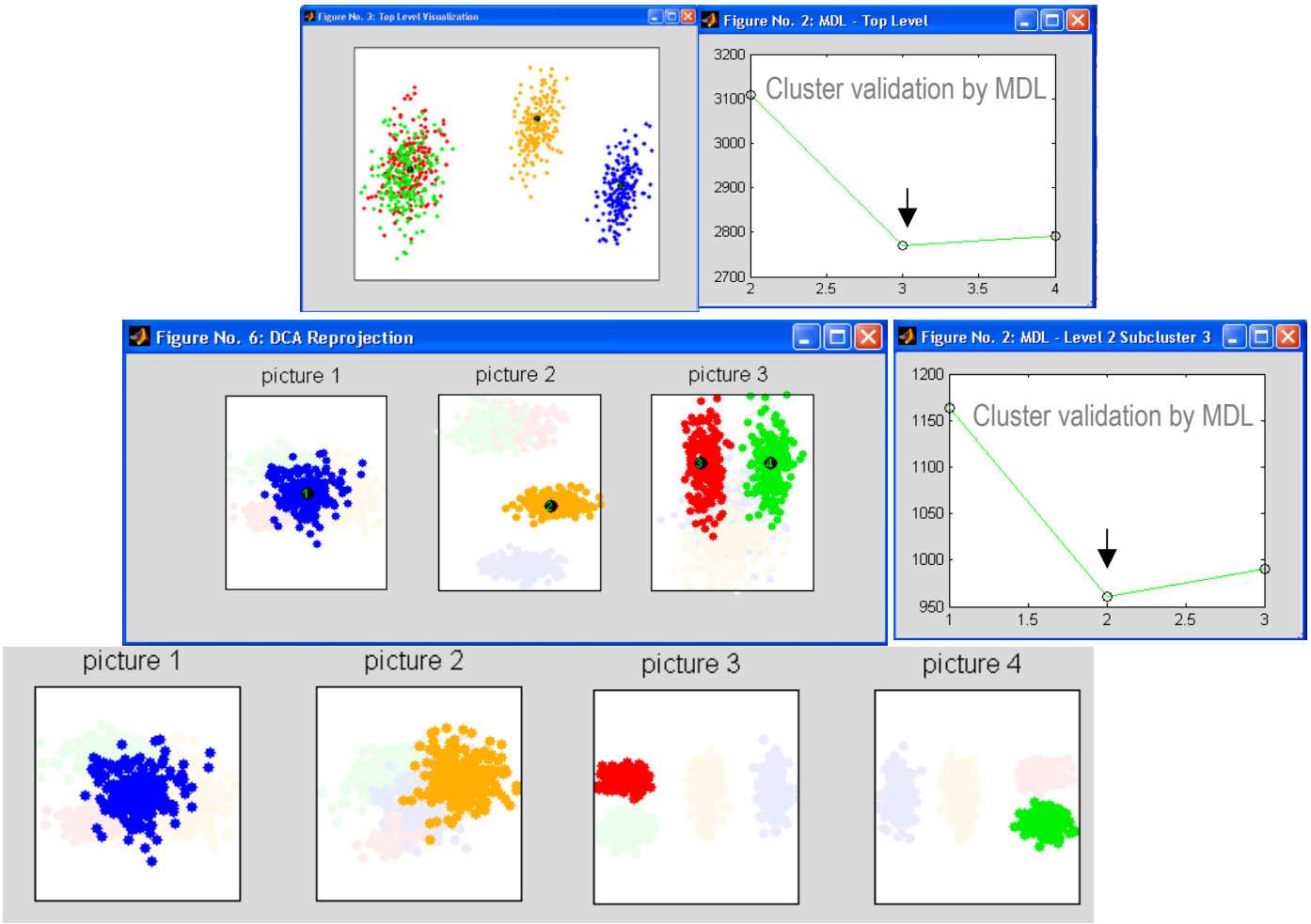
Visual Statistical Data Analyzer -- VISDA

Developer Center -- POC	Georgetown – Joseph Wang
Adopter Center -- POC	Wistar – Louise Showe
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

VISDA Project Objectives

- Develop a Functional Requirements and Design Specification document, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code for the following functionality:
  - Cluster modeling using hierarchical mixture modeling
  - Dimension reduction by principle component analysis (PCA), discriminatory component analysis (DCA) and project pursuit method (PPM)
  - Cluster formation by soft data clustering using expectation-maximization (EM) algorithm
  - Cluster validation by minimum description length (MDL) criterion
  - Cluster visualization by hierarchical cluster display
  - Graphical user interface (GUI) for VISDA set-up, data analysis, and data visualization
- Execute on Test Approach

“Simulation with Truth”



SIG 2:
Genome Annotation
(6, 41, 10-15)

The mission of the Genome Annotation SIG is to provide data and tools that will greatly enhance the cancer research community's access to high quality, comprehensive gene annotations. Having standardized access to these data sources will support a consistent view of all available gene information. This will be achieved by adapting existing software that meets the needs of the user community to comply with caBIG and by creating new software and tools.

Cancer Molecular Pages

Developer Center -- POC	Burnham – Kutbuddin Doctor
Adopter Center -- POC	TBD
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

Cancer Molecular Pages Project Objectives

- Develop use cases for Cancer Molecular Pages
- Develop Functional Requirements and Design Specification documents, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Create a prototype of the Cancer Molecular Pages application and deploy at Adopter site
- Execute on Test Approach
- Install caArray and migrate legacy data to the system

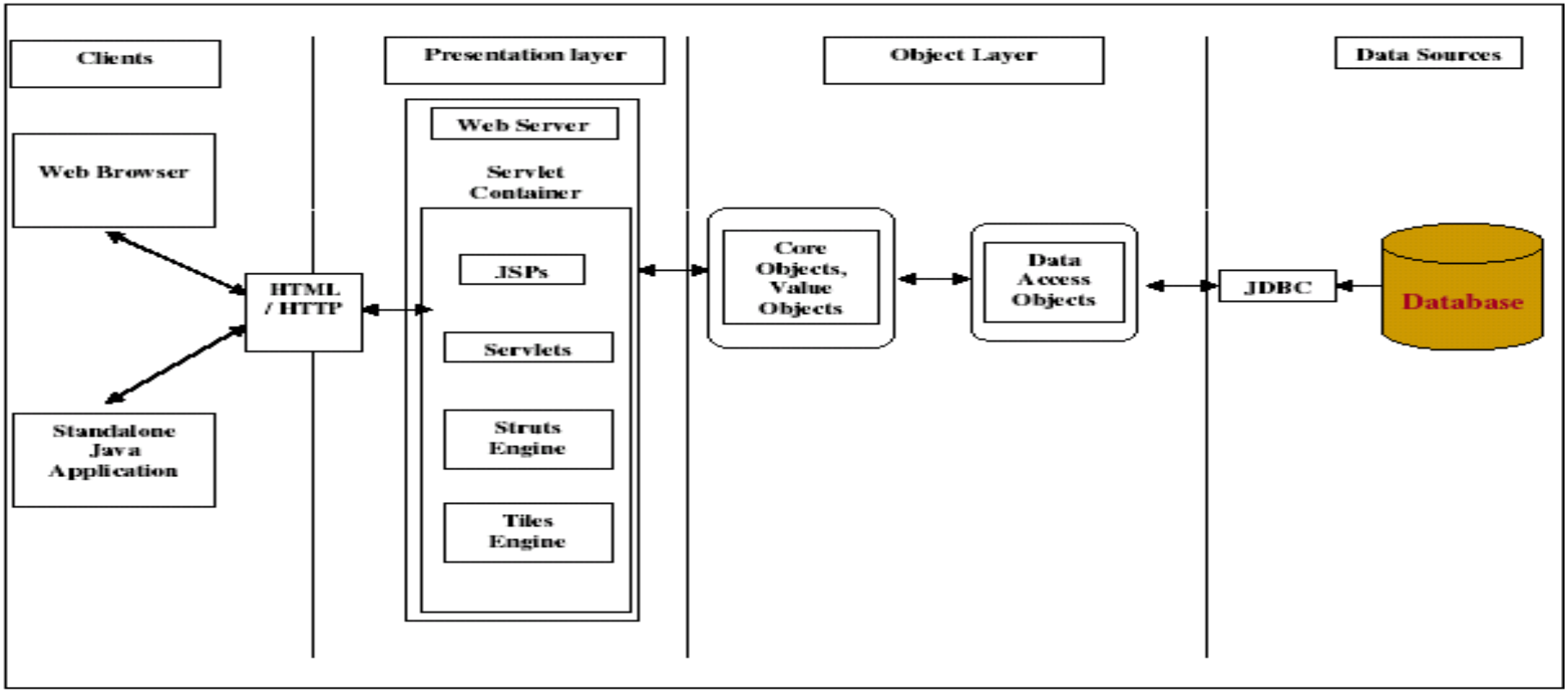
FunctionExpress

Developer Center -- POC	Wash U – Rakesh Nagarajan
Adopter Center -- POC	Wistar – Harold Riethman
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

FunctionExpress Project Objectives

- Develop a Use Case document, in collaboration with the Adopter Center(s)
- Develop a Functional Requirements collaboration with the Adopter Center(s)
- Develop a Design Specification document and UML class diagram of the system.
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code to achieve the following goals:
  - Implementation of the object and data models for the web application
  - Implement web application for annotation-basedgene search and for display of gene literature networks .
  - Import microarray data from external sources, including caArray
  - Import annotation data from external sources
- Execute on Test Approach
- Deploy beta system to Adopter site

Architecture Diagram of FE



High Level Architecture Diagram of Web Application

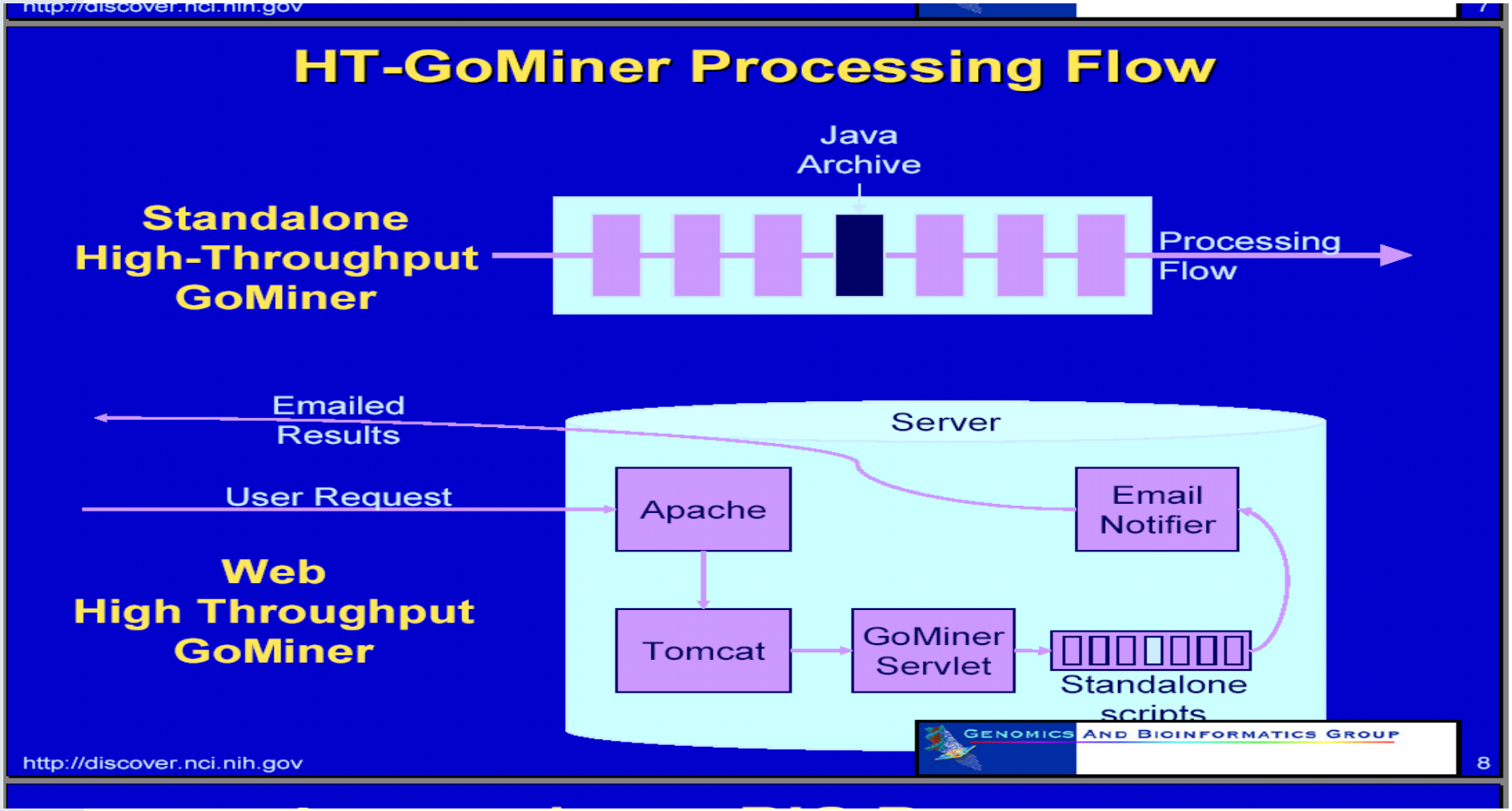
GoMiner

Developer Center -- POC	NCI-CCR – David Kane
Adopter Center -- POC	Wistar – Harold Riethman
caBIG Compliance Level to be Achieved	Silver
Project duration	8 months

GoMiner Project Objectives

- Create a use case document for GoMiner, including the use of BioCarta for classification
- Develop Functional Requirements and Design Specification documents, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces.
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code to achieve the following milestones:
  - Implement an XML-based output format with explicit meta-data descriptors
  - Implement a mechanism to retrieve data from caArray with a mechanism to specify a desired data set
  - Develop a web services API for GoMiner
- Execute on Test Approach
- Package and release the GoMiner source code under an Open Source license

High-Throughput GoMiner Processing Flow



HapMap, Vertebrate Promoter DB

Developer Center -- POC	Cold Spring Harbor – Brian Gilman
Adopter Center -- POC	Wistar – Harold Riethman MSKCC- Alex Lash
caBIG Compliance Level to be Achieved	Silver
Project duration	9 months

HapMap, Vertebrate Promoter DB Project Objectives

- Develop a Functional Requirements and Design Specification document, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code to achieve the following milestones:
  - Establish use cases for the two databases
  - Serve HapMap and the VPD by the DAS database
  - Import HapMap/VPD Data into the caBIO database
  - Augement caBIO system to serve Hapmap and VPD data
  - Validate import of HapMap and VPD data to caBIO object model
- Execute on Test Approach

Protein Information Resource (PIR)

Developer Center -- POC	Georgetown -- Cathy Wu
Adopter Center -- POC	Penn – David Fenstermacher
caBIG Compliance Level to be Achieved	Gold
Project duration	7 months (for first phase of project)

PIR Project Objectives

- In collaboration with the Adopter center, develop use cases for grid-enabled PIR
- Develop Functional Requirements and Design Specification documents, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Provide the following to support a PIR middleware layer:
  - Database to object mapping description
  - Produce a web services layer to access data
  - Publish description of service in WSDL with XML schema that defines input and output objects
- Execute on Test Approach

UniProt Report

PIR View

UniProt Entry: P00439

ENTRY INFORMATION	
ENTRY NAME	PH4H_HUMAN
ACCESSION NUMBERS	P00439; Q16717; Q8TC14
CREATED	Release 01, 21-JUL-1986
SEQUENCE UPDATE	Release 01, 21-JUL-1986
ANNOTATION UPDATE	Release 44, 05-JUL-2004

NAME AND ORIGIN OF THE PROTEIN	
PROTEIN NAME	Phenylalanine-4-hydroxylase
DESCRIPTION	(EC 1.14.16.1) . PAH; Phe-4-monooxygenase
GENE NAME	PAH
SOURCE ORGANISM	Homo sapiens
TAXONOMY ID	9606 [NCBI, NEWT]
LINEAGE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mam

REFERENCES	
[1]	Kwok SCM; Ledley FD; Dilella AG; Robson KJH; Woo SLC **Nucleotide sequence of a full-length complementary DNA clone an** 1985, *Biochemistry*, 24, 556-561 *Position*: SEQUENCE FROM N.A. *Comments*: tissue=Liver PubMed: 2986678; Medline: 85199778;

COMMENTS	
CATALYTIC ACTIVITY	L-phenylalanine + tetrahydrobiopterin + O(2) = L-tyrosine + 4a-hydroxytetrahydrobio
COFACTOR	Ferrous ion.
ENZYME REGULATION	N-terminal region of PAH is thought to contain allosteric binding sites for phenylalanine and to constitute an "inhibitory" regulates the activity of a catalytic domain in the C-terminal portion of the molecule.
PATHWAY	Catabolism of phenylalanine; first (rate-limiting) step.
SUBUNIT	Homodimer.
POLYMORPHISM	The Glu-274 variant occurs on approximately 4% of African-American PAH alleles. The enzyme activity of the variant indistinguishable from that of the wild-type form.
DISEASE	Defects in PAH are the cause of phenylketonuria (PKU) [MIM:261600]. PKU is an autosomal recessive inborn error phenylalanine metabolism, due to severe phenylalanine hydroxylase deficiency. It is characterized by blood concentratio phenylalanine persistently above 1200 mumol (normal concentration 100 mumol) which usually causes mental retardatio low phenylalanine diet is introduced early in life). They tend to have light pigmentation, rashes similar to eczema, epileps hyperactivity, psychotic states and an unpleasant "mousy" odor.
DISEASE	Defects in PAH are the cause of non-phenylketonuria hyperphenylalaninemia (Non-PKU HPA) [MIM:261600]. Non- is a mild form of phenylalanine hydroxylase deficiency characterized by phenylalanine levels persistently below 600 mun allows normal intellectual and behavioral development without treatment. Non-PKU HPA is usually caused by the com of a mild hyperphenylalaninemia mutation and a severe one.
DISEASE	Defects in PAH are the cause of hyperphenylalaninemia (HPA) [MIM:261600]. HPA is the mildest form of phenylalani hydroxylase deficiency.
SIMILARITY	Belongs to the biopterin-dependent aromatic amino acid hydroxylase family.
ONLINE INFORMATION	PAHdb; Phenylalanine hydroxylase locus knowledgebase; Belongs to the biopterin-dependent aromatic amino acid hydroxylase family.

DATABASE CROSS-REFERENCES	
EC	1.14.16.1.
EMBL	K03020,AAA60082.1. [GenBank, DDBJ] U49897,AAC51772.1. [GenBank, DDBJ] S61296,AAD13926.1. [GenBank, DDBJ] BC026251,AAH26251.1. [GenBank, DDBJ]
GENEW	HGNC:8582,PAH.
GO	GO:0004505,F:phenylalanin GO:0008652,P:amino acid b
HSC_2DPAGE	P00439,HUMAN.
INTERPRO	IPR001273,Aaa_hydroxyla IPR002912,ACT. IPR005961,Phe4hydrxlaseT
MIM	261600.
PDB	1DMW,2001-03-24. 1J8T,2002-05-22. 1J8U,2002-05-22. 1KW0,2003-01-28. 1LRM,2002-06-12. 1MMK,2003-09-04. 1MMT,2003-09-04. 1PAH,1999-01-13. 2PAH,1999-10-06. 3PAH,1999-04-27. 4PAH,1999-04-27. 5PAH,1999-04-27.
PFAM	
PIR	
PRINTS	
PRODOM	
PROSITE	
REACTOME	
TIGRFAMS	

KEYWORDS

Oxidoreductase; Monooxygenase; Allosteric enzyme; Phenylketonuria; Phosphorylation; Phenylalanine catabolism; Iron; Disease mutation; Polymorphism; 3D-structure

FEATURES

Feature	Description	Begin Position	End Position	Length
MOD_RES	Phosphoserine (by PKA) (BY SIMILARITY)	16	16	1
METAL	Iron (BY SIMILARITY)	285	285	1
METAL	Iron (BY SIMILARITY)	290	290	1
METAL	Iron (BY SIMILARITY)	330	330	1
VARIANT	S -> P (in PKU) /FTId=VAR_000869	16	16	1
VARIANT	Q -> L (in HPA) /FTId=VAR_009239	20	20	1
VARIANT	F -> L (in PKU; haplotype 1) /FTId=VAR_000870	39	39	1
VARIANT	(in PKU; haplotypes 9,21) /FTId=VAR_000871	39	39	1
VARIANT	S -> L (in PKU) /FTId=VAR_000872	40	40	1
VARIANT	L -> F (in PKU) /FTId=VAR_000873	41	41	1
VARIANT	L -> P (in PKU; mild) /FTId=VAR_009240	41	41	1

ADDITIONAL INFORMATION FROM iProClass — Go to iProClass

CROSS-REFERENCE	RefSeq: NP_000268 phenylalanine hydroxylase LocusLink: 5053 phenylalanine hydroxylase(PAH) NCBI GI#: 2462722; 189937; 4557819
PIRSF FAMILY	PIRSF000336: phenylalanine 4-monooxygenase
GENE ONTOLOGY	*Molecular Function* GO:0004505: phenylalanine 4-monooxygenase activity [INTERPRO; evidence:IEA] [SPEC; evidence:IEA] [PMID:3856322; evidence:TAS] GO:0016597: amino acid binding [INTERPRO; evidence:IEA] GO:0005506: iron ion binding [INTERPRO; evidence:IEA] GO:0004497: monooxygenase activity [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] GO:0003824: catalytic activity [SPKW; evidence:IEA] GO:0016491: oxidoreductase activity [SPKW; evidence:IEA] *Biological Process* GO:0009072: aromatic amino acid family metabolism [INTERPRO; evidence:IEA] GO:0008652: amino acid biosynthesis [PMID:3856322; evidence:TAS] GO:0008152: metabolism [INTERPRO; evidence:IEA] GO:0006559: phenylalanine catabolism [INTERPRO; evidence:IEA] [SPKW; evidence:IEA] [UniProt:P00439; evidence:none]
ENZYME/FUNCTION	EC 1.14.16.1 EC-IUBMB , KEGG, BRENDA, WIT, MetaCyc *Nomenclature:* Oxidoreductases; Acting on paired donors, with incorporation or reduction of molecular oxygen; With reduced pteridine as one donor, and incorporation of one atom of oxygen; phenylalanine 4-monooxygenase *Reaction:* L-phenylalanine + tetrahydrobiopterin + O_2 = L-tyrosine + 4a-hydroxytetrahydrobiopterin
PATHWAY	KEGG: Metabolism; Amino Acid Metabolism; Phenylalanine, tyrosine and tryptophan biosynthesis [PATH:hsa00400].
STRUCTURE	PDB: 1PAH(117-424,100.0%) ; 1DMW:A(118-424,100.0%) ; More 1DMW: SCOP CATH FSSP MMDB PDBsum 1IN9: SCOP CATH FSSP MMDB PDBsum More

http://www.pir.uniprot.org/cgi-bin/upEntry?id=PH4H_HUMAN

caBIG cancer Biomedical Informatics Grid

SEED

Developer Center -- POC	U of Chicago – Ed Frank
Adopter Center -- POC	Georgetown – Cathy Wu
caBIG Compliance Level to be Achieved	Silver
Project duration	TBD

SEED – For subsystem definition

missing gene in human

Human functional variant

Organism	Variant Code	ASPDC	KPHMT	KPRED	KARED	PBAL	PANF	PANK	PPCS	PPCDC	PPAT	DPCK
Mycoplasma genitalium [B]	1											272
Streptococcus pneumoniae R6 [B]	1							741	1109	1110	1781	873
Homo sapiens [E]	1						18781,24457,3886,421,5247	26330	8649	20554	4329	4329,4427,4428
Methanocaldococcus jannaschii [A]	0				1644				1001	1001	1117	
Haemophilus influenzae Rd KW20 [B]	1				652		941	601	920	920	621	857
Aeropyrum pernix [A]	0		517						1396	1396		
Pyrobaculum aerophilum str. IM2 [A]	0		2402						831	831	608	
Buchnera aphidicola str. APS (Acyrthosiphon pisum) [B]	1		196		566	195					554	202
Synechocystis sp. PCC 6803 [B]	1		2844		2956	1587	1768	1542	1366	1366	2569	2904
Saccharomyces cerevisiae [E]	0		3766	1690	3279	5013		2331	5073	2556,5673,5824	4971	2011
Bacillus anthracis str. A2012 [B]	1	2391		2694	2247,2663		4407,4496	3725,963	4786	4786	4914	5560
Helicobacter pylori J99 [B]	1	30	364		310	6		792	775	775	1364	766
Thermotoga maritima [B]	1	931	1710		544	1067		875	1671	1671	734	1372
Mycobacterium tuberculosis H37Rv [B]	0	3604	2227	2575		3605		1094,3603	1393	1393	2967	1633
Bacillus subtilis subsp. subtilis str. 168 [B]	1	2245	2247	1446	2832	2246	323,667	2381	1572	1572	1504	2909
Escherichia coli K12 [B]	1	131	134	422	3704	133	3200	3890	3575	3575	3570	103

SIG 3:
Microarray Repositories
(3, 47, 15)

The mission of the Microarray Repositories SIG is to identify and prioritize the needs of the larger cancer research community with respect to the capture, storage, and utilization of microarray data and related types of genetic data. Specifically, we will address the need for:

• A database for the storage and retrieval of microarray data and related data types that can be incorporated into the larger scheme of federated databases that store clinically relevant data and other relevant data types.

• Software that facilitates the capture of important microarray experimental information and automatically loads it into a database.

• Software that facilitates the querying of microarray databases and the retrieval of data.

• Consumers and producers of microarray data to readily exchange data

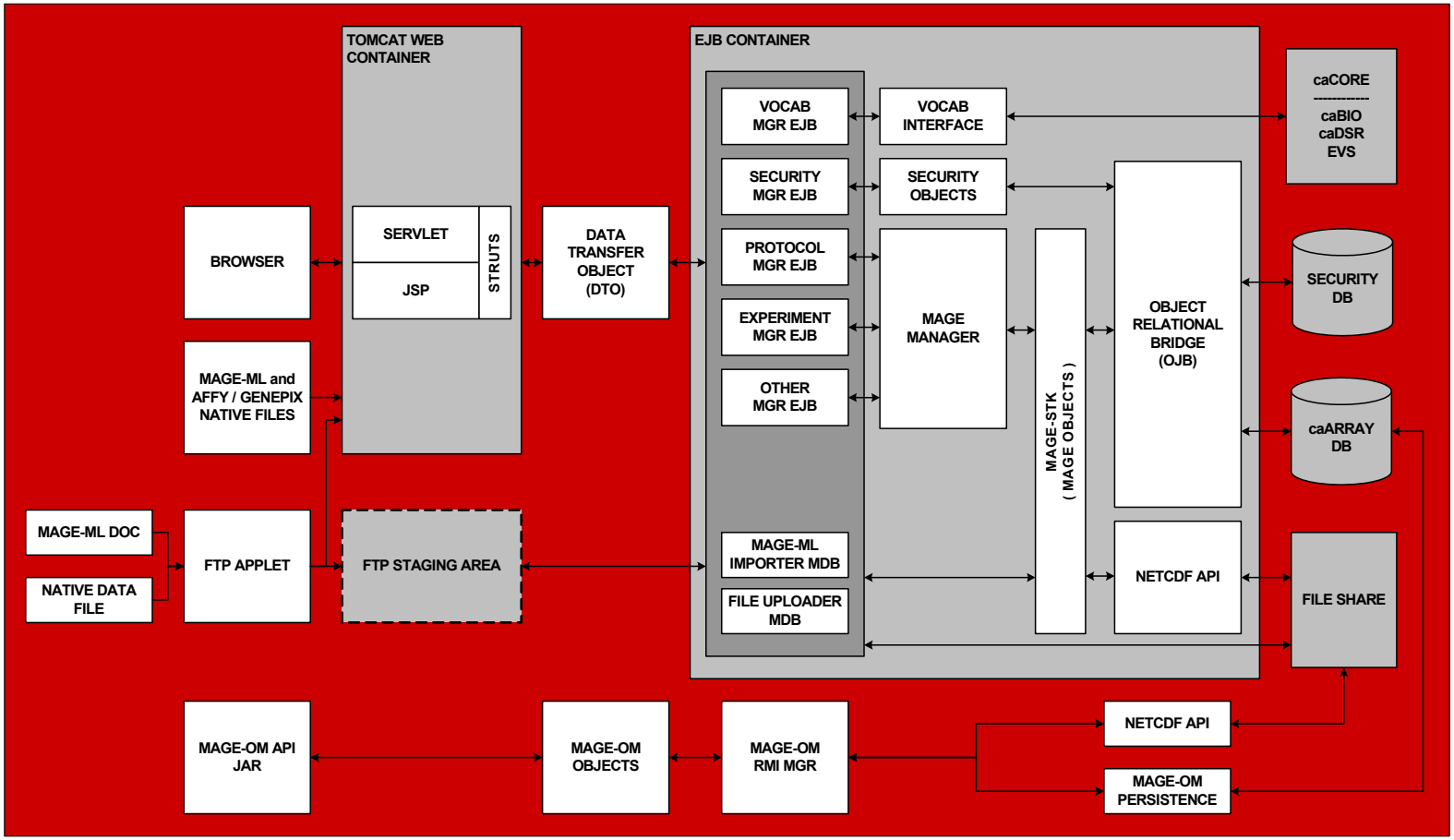
caArray

Developer Center -- POC	NCICB – Mervi Heiskanen
Adopter Center -- POC	Wistar – Louise Showe NYU – Judith Goldberg Georgetown – Arnie Miles
caBIG Compliance Level to be Achieved	N/A
Project duration	7 months

caArray Project Objectives

- Create a Test Procedures document
- Install the caArray software and create an Installation Guide
- Test the experiment annotation, data loading and data transfer functionality of caArray using test data provided by NCICB and local data
- Review and update the Usage Guide provided by NCICB
- Train users; Review and update the training materials provided by NCICB

caArray Architecture



NCI-60 Data Sharing

Developer Center -- POC	NCI-CCR – David Kane
Adopter Center -- POC	MSKCC – Alex Lash
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

NCI-60 Project Objectives

- Create a Work Plan for the project
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures that data are transformed accurately
- Create the following data resources
  - XML-based representation of NCI-60 cell line characterization
  - MAGE-ML files and source data files for Affymetrix U95 and U133 NCI-60 datasets
  - MAGE-ML toxicology files for two NCI-60 drug screen datasets
  - MAGE-ML files and source data files for the NCI-60 9,706 clone cDNA gene expression data sets
  - SKYWEB files for the NCI-60 karyotyping data
  - MAGE-ML files and source data files for the aCGH NCI-60 datasets
- As appropriate and feasible, upload these files to the current caBIG reference implementation for these standards, e.g. caArray.
- Execute on Test Approach

Zebrafish Microarray Data Sharing

Developer Center -- POC	Thomas Jefferson – Jack London
Adopter Center -- POC	MSKCC – Alex Lash
caBIG Compliance Level to be Achieved	Silver
Project duration	TBD

SIG 4:
Pathways
(3, 35, 10)

The Pathway SIG endeavors to support basic research and ICR tool development by helping to provide the cancer research community with easy access to pathway data and commonly used pathway analysis tools.

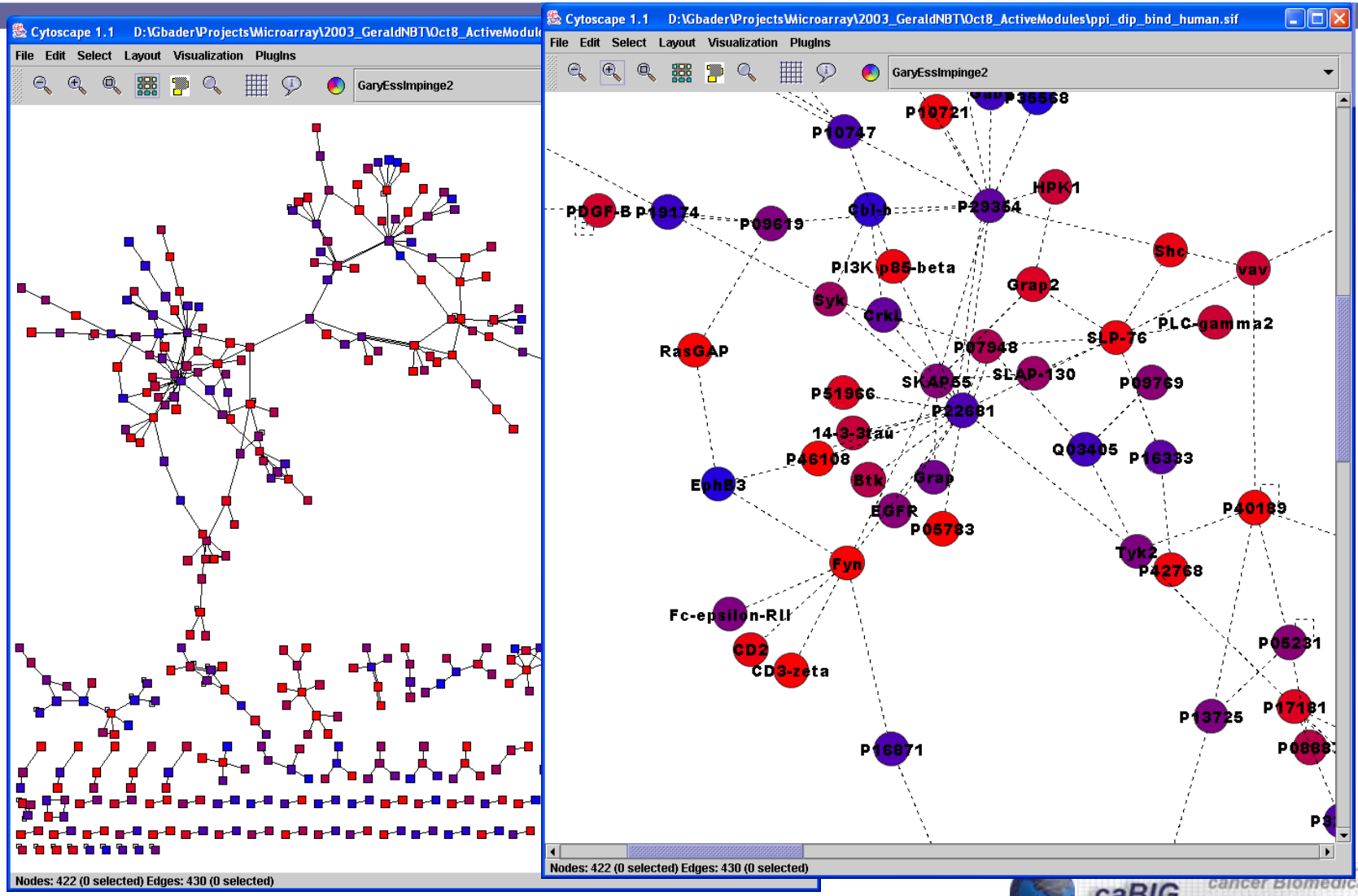
Pathways Tool Development

Developer Center -- POC	MSKCC – Gary Bader
Adopter Center -- POC	OHSU – Shannon McWeeney
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

Pathways Project Objectives

- Develop a Functional Requirements and Design Specification document, in collaboration with the Adopter Center and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Write code to achieve the following milestones:
  - Build I/O functionality for BioPAX format in the Data Services Framework used by cPath and Cytoscape
  - Expand cPath data model to BioPAX 1.0
  - Update Cytoscape to show a simple view of BioPAX data
  - Create Cytoscape plugin for loading caBIG gene annotation and expression data
- Execute on Test Approach

Cytoscape Screen Shots



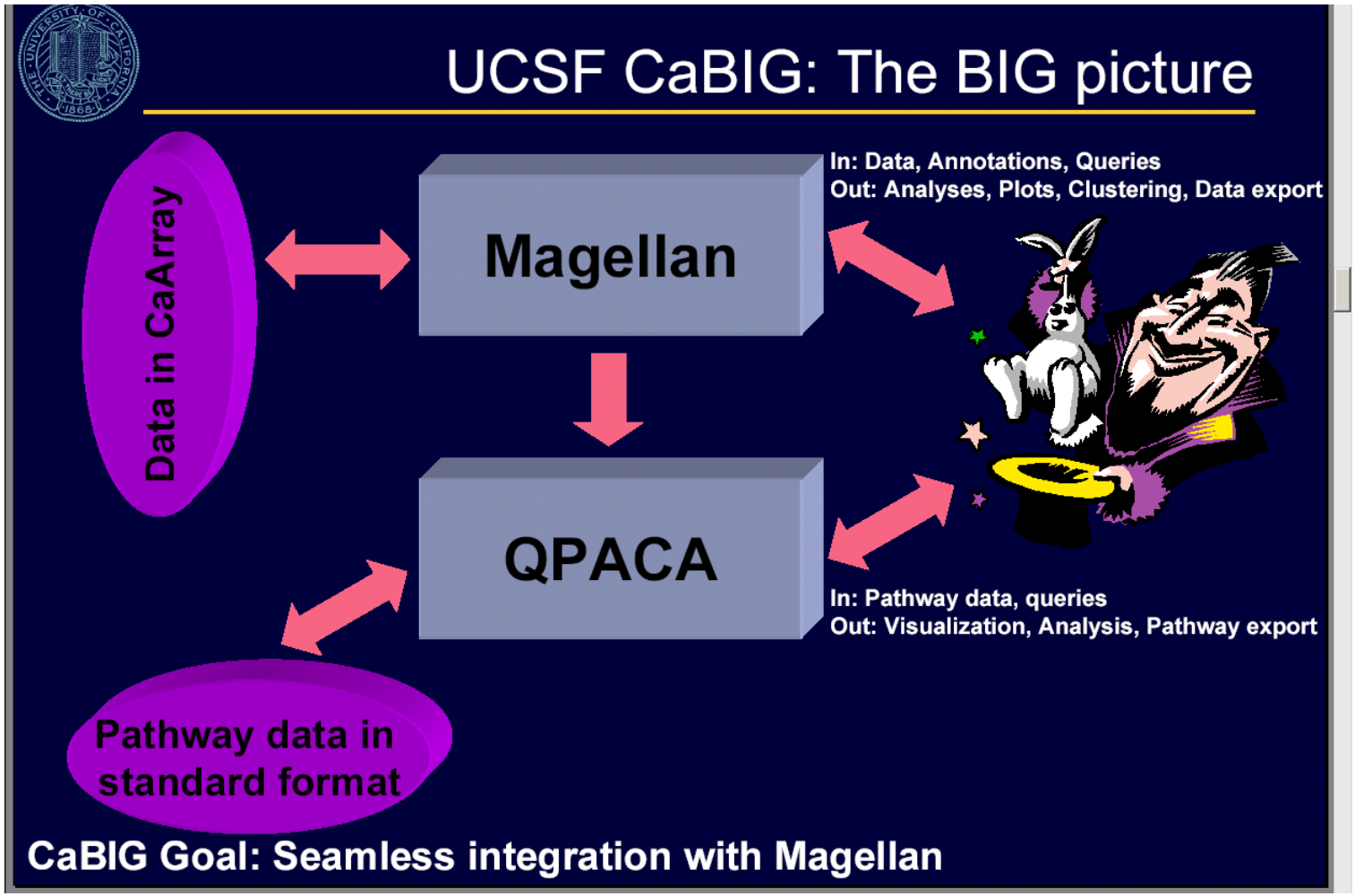
Quantitative Analysis of Pathways in Cancer (QPACA)

Developer Center -- POC	MSKCC – Gary Bader
Adopter Center -- POC	OHSU – Shannon McWeeney
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

QPACA Project Objectives

- Develop a Functional Requirements and Design Specification document, in collaboration with the Adopter Center(s) and the cross-cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Make current version of QPACA available to the Adopter for evaluation
- Write code to achieve the following milestones:
  - Make QPACA interoperable with Magellan
  - Refine QPACA statistical methods
  - Make QPACA interoperable with caBIG-defined Pathways Exchange Standards
- Execute on Test Approach

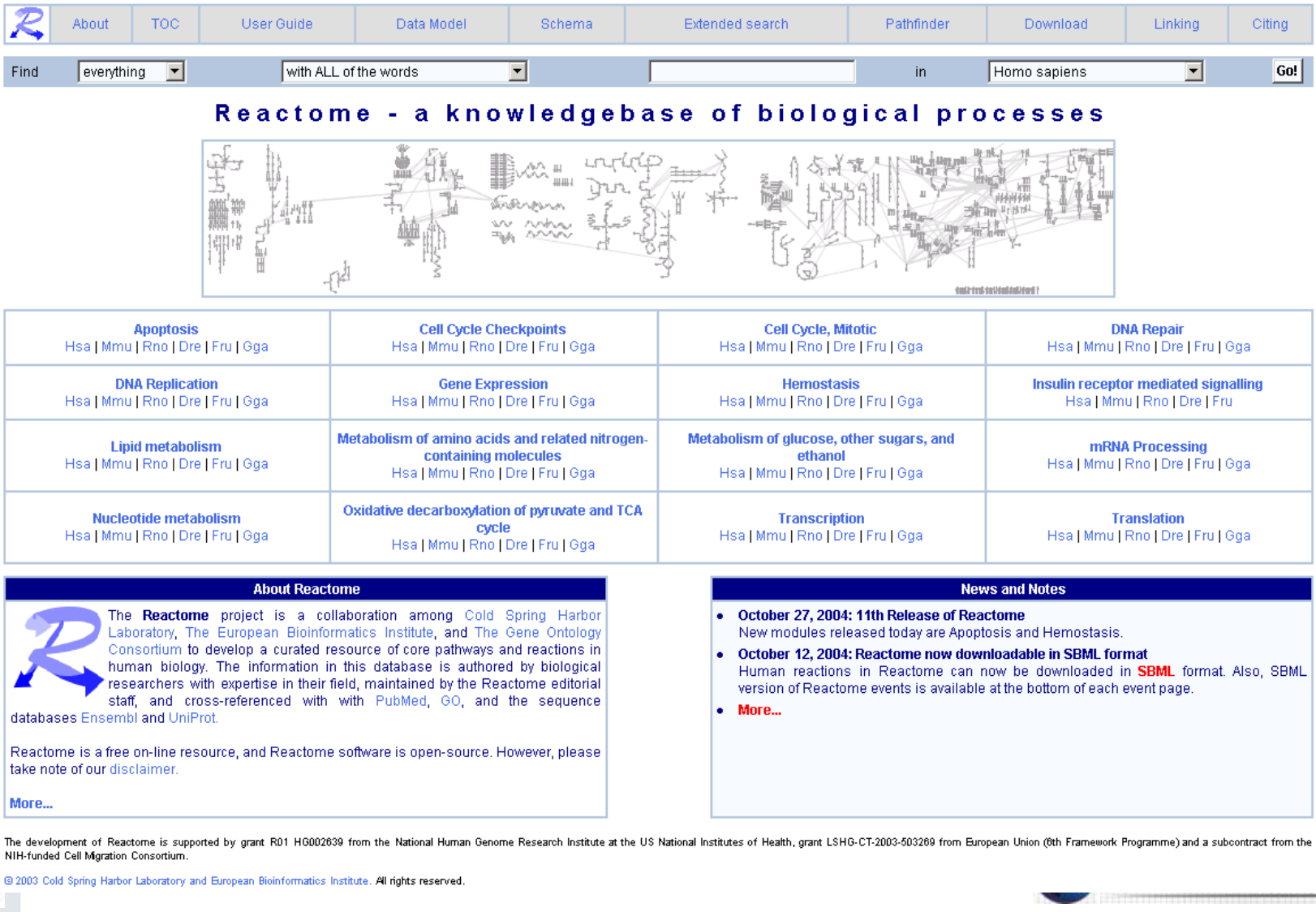
QPACA and caBIG



Reactome Project

Developer Center -- POC	Cold Spring Harbor – Brian Gilman
Adopter Center -- POC	MSKCC – Gary Bader
caBIG Compliance Level to be Achieved	Silver
Project duration	TBD

Reactome Home Page



SIG 5/6: Proteomics (3, 43, 10)

The Proteomics Tools SIG is focused on:

- Tools and technologies which are necessary for cancer centers to store, annotate, and analyze the growing proteomics data sets
- Providing flexible tools for data and metadata storage, so that emerging technologies can be incorporated into developed systems
- Integrating proteomics data with other data through use of appropriate CDEs and architectures

Proteomics LIMS

Developer Center -- POC	Fox Chase – Michael Ochs
Adopter Center -- POC	Moffitt – Steven Eschrich
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

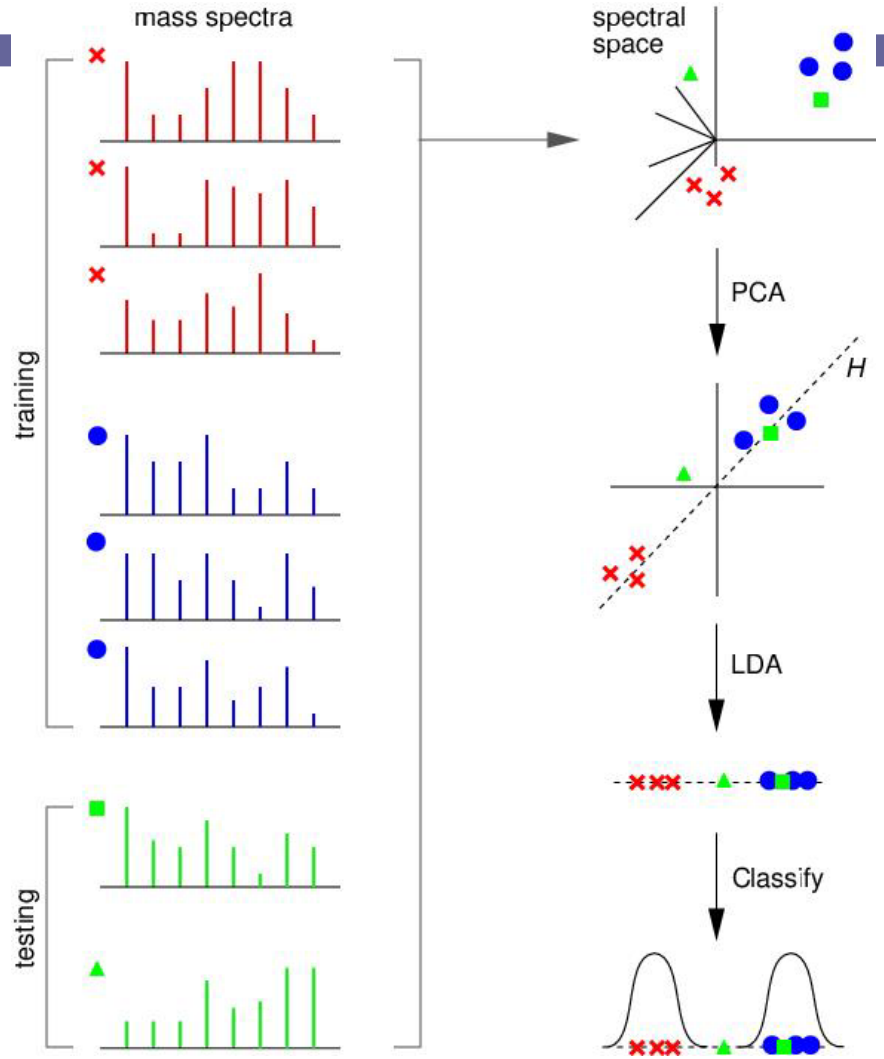
Proteomics LIMS Project Objectives

- Develop use cases to represent the lab processes around 2D gel electrophoresis
- Develop a Functional Requirements and Design Specification document, in collaboration with the Adopter Center and the cross-cutting Workspaces
- Create a UML model to represent the system
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Create mock-ups of the system user interfaces for review by the Adopter
- Execute on Test Approach
- Deploy prototype system to the Developer and Adopter sites

Q5

Developer Center -- POC	Dartmouth – David Jewell
Adopter Center -- POC	OHSU – Shannon McWeeney
caBIG Compliance Level to be Achieved	Silver
Project duration	TBD

Q5: Disease Classification by Mass Spec Pattern Recognition



Proteome Analysis

Expression analysis: proteins (mass spectrometry)

Algorithm uses PCA followed by LDA

probabilistic classification of healthy vs. disease whole serum samples using mass spectrometry

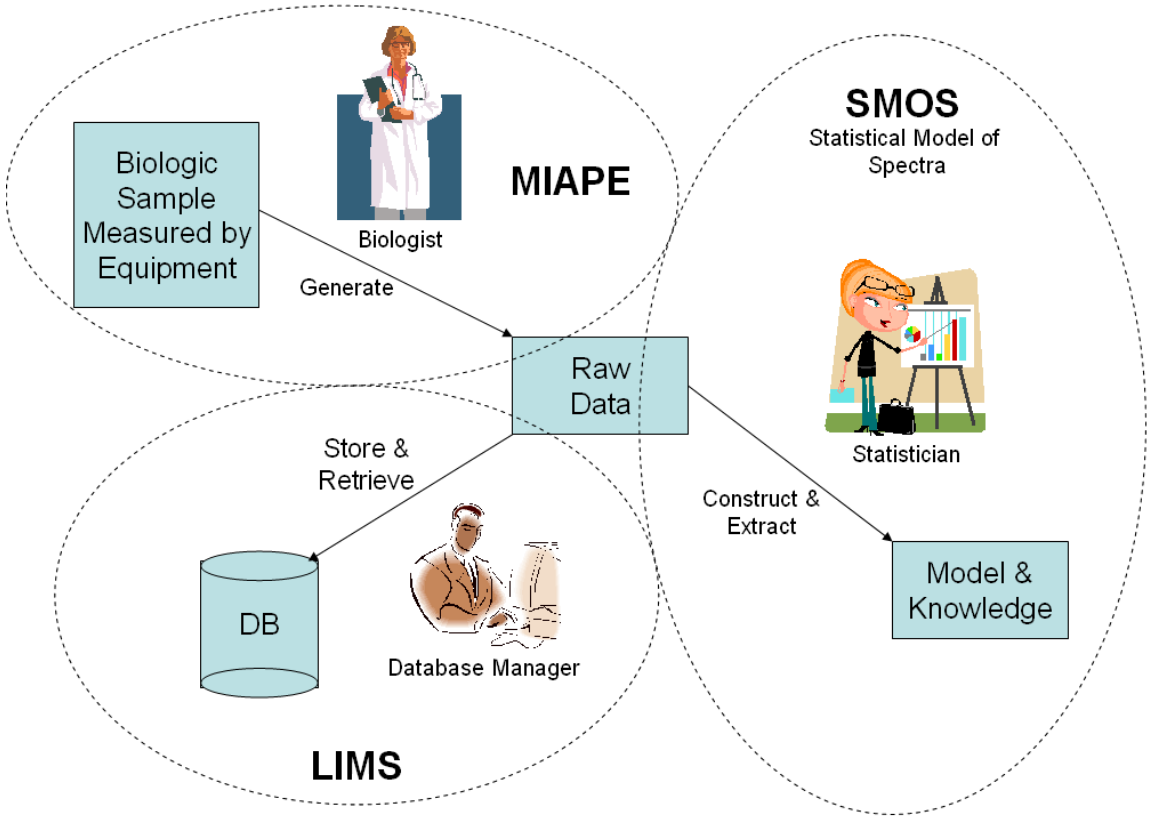
RProteomics

Developer Center -- POC	Duke – Patrick McConnell
Adopter Center -- POC	OHSU – Shannon McWeeney Penn – David Fenstermacher
caBIG Compliance Level to be Achieved	Gold
Project duration	9 months

RProteomics Project Objectives

- In collaboration with the Adopter centers, develop use cases for RProteomics
- Develop Functional Requirements and Specification document, in collaboration with the Adopter Centers and the Cross-Cutting Workspaces
- Create a Risk Management Matrix for the project
- Document a Test Approach that ensures requirements are met
- Describe the RProteomics data structures in UML and as common data elements
- Create source code for
  - R and Java data structures,
  - Java-to-R bridge
  - Java and R grid wrappers
  - Client application
- Document best practices for integrating Java and R
- Present project to ICR and Architecture Workspaces

Relationship between MIAPE and SMOS



SIG 6/6: Translational (1, 40, 10)

The Translational Tools Special Interest Group is focused on:

- Tools and technologies which are necessary for cancer centers to integrate clinical data with experimental data, as well as experimental design tools that provide assistance to biomedical investigators embarking on translational research methodology
- Assuring that clinical data and studies more effectively utilize genomics and proteomics research data in cancer research and patient care
- Creating guidelines to aid in the design of experimental studies

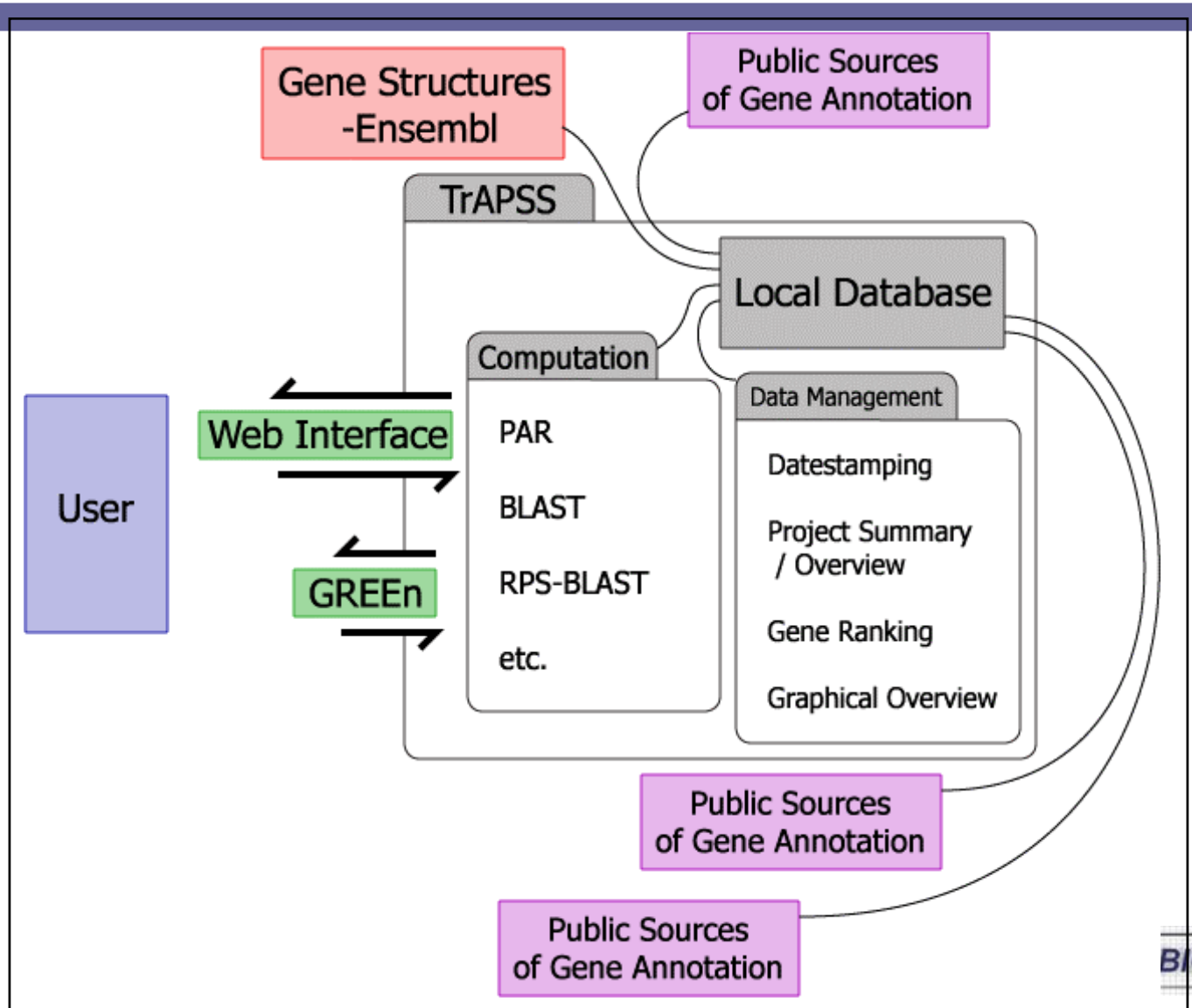
Transcript Annotation Prioritization and Screening System (TrAPSS)

Developer Center -- POC	U of Iowa – Terry Braun
Adopter Center -- POC	Wistar – Harold Riethman
caBIG Compliance Level to be Achieved	Silver
Project duration	12 months

TrAPSS Project Objectives

- Create a Risk Management Matrix for the project
- Make current version of TrAPSS available to the Adopter and train them on the use of the system
- Develop a Use Case document, in collaboration with the Adopter Center
- Develop a Functional Requirements and Specification Document in collaboration with the Adopter Center and the cross-cutting Workspaces
- Document a Test Approach that ensures requirements are met
- Document and implement necessary changes to the TrAPSS database
- Implement the changes necessary to the TrAPSS modules as defined in the Requirements and Specification Document
- Execute on Test Approach
- Deploy system to Adopter site
- Create a Developer's guide

TrAPSS System



ICR Workspace at a Glance
- Conclusion::Goals Defined (drafted)

- One Year Goals:
  - The majority of projects will target Silver Level compliance, as outlined in the caBIG Compatibility Guidelines document. A few projects will be selected as reference implementations of the Grid architecture. All projects will comply with the caBIG principles of open source, open access, open development and federation.
- Three year goals:
  - 1. Bring more selected data sources and applications into Gold-level compliance, i.e. make them grid resources.
  - 2. Show how a federated set of resources that are syntactically and semantically compatible can be used to perform powerful analyses across the cancer research community.
- Five year Goals
  - Connect the ICR data and applications to resources in the Tissue Banks and Pathology Tools Workspace as well as the Clinical Trials Management System Workspace

ICR Workspace at a Glance
- Conclusion::Perceived Challenges

- "Tightening" the group – how do projects "fit" together?
- "Cementing" the Developer-Adopter relationships (More use-cases need to be developed jointly)
- Increasing awareness crossing…
  - Workspaces (beyond liaisons)
  - SIGs (distinction?/commonality?)
- Uniform training
- Uniform adoption of caBIG Archigecture, CDEs and practices
- Balancing Software Engineering with Science
  - Discussions
  - Effort
- A "Vision" for the "Grid"